

# Sleeping newborns extract prosody from continuous speech

Citation for published version (APA):

Sambeth, A., Ruohio, K., Alku, P., Fellman, V., & Huotilainen, M. (2008). Sleeping newborns extract prosody from continuous speech. *Clinical Neurophysiology*, 119(2), 332-341. <https://doi.org/10.1016/j.clinph.2007.09.144>

## Document status and date:

Published: 01/01/2008

## DOI:

[10.1016/j.clinph.2007.09.144](https://doi.org/10.1016/j.clinph.2007.09.144)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# Sleeping newborns extract prosody from continuous speech

Anke Sambeth<sup>a,b</sup>, Katja Ruohio<sup>a</sup>, Paavo Alku<sup>c</sup>, Vineta Fellman<sup>d,e</sup>,  
Minna Huotilainen<sup>a,f,\*</sup>

<sup>a</sup> Brain and Music Team, Cognitive Brain Research Unit, Department of Psychology, P.O. Box 9, FIN-00014, University of Helsinki, Finland

<sup>b</sup> Department of Neuropsychology and Psychopharmacology, Faculty of Psychology, Maastricht University, The Netherlands

<sup>c</sup> Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Espoo, Finland

<sup>d</sup> Department of Pediatrics, University of Lund, Sweden

<sup>e</sup> Hospital for Children and Adolescents and Department of Obstetrics and Gynaecology, Helsinki University Central Hospital, Helsinki, Finland

<sup>f</sup> Helsinki Collegium for Advanced Studies, University of Helsinki, Finland

Accepted 23 September 2007

## Abstract

**Objective:** Behavioral experiments show that infants use both prosodic and statistical cues in acquiring language. However, it is not yet clear whether these prosodic and statistical tools are already present at birth.

**Methods:** We recorded brain responses of sleeping newborns to natural sounds rich in prosody, namely singing and continuous speech, and to two impoverished manipulations of speech. A total of 11 newborns were presented with continuous speech, singing, and degraded speech, while MEG was recorded.

**Results:** We found that a brain response elicited to the prosodically rich singing and continuous natural speech conditions decreased dramatically when the prosody in the speech was impoverished.

**Conclusions:** We claim that this response is the indicator of the infants' sensitivity to prosodic cues in language, which is already present at birth during natural sleep.

**Significance:** The indicators of detection of prosody may be crucial in assessing the normal and abnormal cortical function in newborns, especially of those infants at-risk for language problems.

© 2007 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.

**Keywords:** Continuous speech; Degraded speech; Magnetoencephalography; Newborn; Prosody

## 1. Introduction

In order to acquire their native language, infants must learn to identify words from continuous streams of speech sounds. Whereas the word boundaries are very clear in written language, they are practically absent in spoken language. How does an infant know when a word has ended and a new began?

During the last decade, numerous studies have examined behaviorally how infants segment speech and when

language acquisition starts. The most common test to study infant language acquisition around the age of 6–12 months is the head turning procedure (for a review, see Kuhl, 2004). Before the test starts, infants are first familiarized to repetitions of the word cues of interest. Subsequently, they are presented with fragments of speech that either contain the newly learned stimuli or not. One trial works as follows. When an infant turns the head to look at a light from the side, stimulus presentation starts. When the infant stops watching and turns the head back again, the stimulation stops. The length of time that the infant has looked into the direction where the sound came from is determined and differences between looking time provide information as to which piece of speech the infant favors.

\* Corresponding author. Address: Cognitive Brain Research Unit, Department of Psychology, P.O. Box 9, FIN-00014, University of Helsinki, Helsinki, Finland. Tel.: +358 9 19129459; fax: +358 9 19129542.  
E-mail address: minna.huotilainen@helsinki.fi (M. Huotilainen).

It has been shown that infants of less than a year old use at least two features in language to segment words (for review, see Kuhl, 2004). First, infants are sensitive to the sequential probabilities between neighboring syllables. In other words, they use the statistical likelihood that a certain syllable follows another syllable. The second important feature in segmenting speech is prosody. Prosodic stress cues of a syllable help infants to identify word boundaries in continuous speech. This study follows the most widely used interpretation of terminology in which prosody is defined as: a suprasegmental feature of speech that is conveyed by fundamental frequency, intensity, and duration.

Saffran et al. (1996) showed that, by the age of 8 months, infants are able to use statistical cues to segment words from fluent speech. Jusczyk et al. (1999) demonstrated that 7.5-month-old infants were able to detect strong/weak stress patterns from fluent speech. In other words, they were able to correctly identify the word onsets using prosody.

Some controversy remains about the importance of prosodic and statistical cues at different ages. In the study of Jusczyk et al. (1999), the 7.5-month-old infants tended to treat the strong syllables as markers for the word onsets in both strong/weak and weak/strong patterned words. They relied heavily on the stress cues, which made it impossible for them to correctly segment the words in the weak/strong condition. Johnson and Jusczyk (2001), furthermore, showed that infants weighed stress cues more heavily than statistical cues at an age of 8 months, although they were also able to segment the speech cues on the basis of statistics alone. Thiessen and Saffran (2003) examined the infant's attention to conflicting cues, presenting speech stimuli in which prosodic and statistical cues gave conflicting information within one and the same word. They found that 9-month-olds used syllable stress as a cue to segment the speech while ignoring the statistical cues. The infants at an age of 7 months, however, attended more to the statistical than to the stress cues.

One reason for this distinction may be the tool used to study language acquisition, namely the head turn procedure. As was pointed out by Saffran (2001), different task factors influence the sort of preference of an infant for novel vs. known stimuli. The length of the delay between familiarization phase and test is important, as is also the number of repetitions of the items during the familiarization phase. These factors were not always similar between the different studies (see Saffran, 2001). Another essential factor is the infant's prior knowledge, since it has been suggested that infants favor speech that best matches with their prior knowledge. Furthermore, attentional capabilities and motor skills may affect the outcome of the experiment.

One objective measure that does not rely on attention and motor skills is neurophysiology, as was also pointed out by Weber et al. (2004). This does not only enable us to study 6- to 12-month-old infants during better con-

trolled situations, but also permits the study of language processing at younger ages when the infants are not yet able to perform the head turning procedure. Event-related potentials (ERP) or fields (ERF) are not only such measures that can be used, but also optical imaging (Pena et al., 2003).

Repetitive auditory stimulation in sleeping human newborns elicits a positive deflection in the ERP and ERF, with a latency of about 200–300 ms (for ERP, see Ellingson et al., 1974; Kurtzberg et al., 1984; Cheour et al., 1998, 2002a; for ERF, see Lengle et al., 2001; Huotilainen et al., 2003; Kujala et al., 2004; Cheour et al., 2004). Kushnerenko et al. (2002b) reported a P350 deflection at birth, together with a small negative peak at 250 ms. Furthermore, they found that these components increased in size until the age of 6–9 months and were still present in 12-month-old infants. Considering the stability of the infant ERP, it may be a good tool to study infant language acquisition at different ages. Furthermore, using optical imaging, Pena et al. (2003) have shown that neonates do respond to language stimuli.

When the repetitive auditory standard stimuli are interspersed with infrequent deviant stimuli, neonates additionally demonstrate change detection and a mismatch response (e.g., Cheour et al., 2002b; Martynova et al., 2003; Huotilainen et al., 2003; Kujala et al., 2004; Sambeth et al., 2006). This response has been associated with the so-called mismatch negativity (MMN), a pre-attentive change-detection response in adults. The MMN is suggested to be evoked by a comparison process between an irregular, deviant stimulus and a sensory memory trace formed by a standard stimulus (e.g., Näätänen and Winkler, 1999). The MMN component is elicited by any discriminable change in repetitive auditory stimulation and it was shown that mismatch responses occur to various kinds of deviances, such as frequency, duration, and even speech deviants already in newborns (Ceponiene et al., 2002; Dehaene-Lambertz and Pena, 1998; Kujala et al., 2004; Pihko et al., 1999; Sambeth et al., 2006). Furthermore, Kushnerenko et al. (2002a) demonstrated that, at the group level, this mismatch response is relatively stable in latency and amplitude from birth to 12 months of age.

Weber et al. (2004) used the mismatch response to study the sensitivity to prosodic cues in two-syllable pseudowords in 4- and 5-month-old infants as well as in adults. They examined the infant's ability to discriminate trochaic (stress/emphasis on first syllable of the word) from iambic (stress on second syllable of the word) stress patterns and vice versa. As expected, the adults displayed a typical mismatch response to both trochaic and iambic items when they were violated. The infants, however, responded in different ways. Whereas no reliable mismatch response could be found in 4-month-old infants to either item, the 5-month-olds displayed significant change detection for the trochaic items. Weber et al. (2004) concluded that 5-month-old infants could separate trochaic items consistently from iambic items. Furthermore, they suggested a

clear development between 4 and 5 months with respect to the processing of different stress patterns relevant for word segmentation.

Although the study of Weber et al. (2004) provided a first valuable step in the study of language acquisition using neurophysiology, one shortcoming of their study is that they used pseudowords in isolated form to test the infant's abilities in word segmentation. More information may be gathered when presenting continuous speech to the infants. We examined infant's responding to word onsets in continuous speech.

Thiessen and Saffran (2003) examined the so-called 'chicken-and-egg' problem: How can infants determine that words in English tend to be stressed on their first syllables if they are not yet familiar with any words? According to Thiessen and Saffran (2003), infants at different ages may use different segmentation strategies. Thiessen and Saffran stated that 7-month-old infants attended more to the statistical than to the stress cues. It should be noted, however, that even though they preferred them less, the infants did also attend to the stress cues. Therefore, the chicken-and-egg problem is still not solved. Do newborns already possess the capability to use prosody?

In the current study, we tested whether the ability to respond to stress cues in continuous speech is present at birth. It must be noted that there are different kinds of prosody, such as linguistic and emotional prosody (Pihan, 2006). Pitch, loudness contours, and speech and articulation rates are mainly related to the linguistic prosody, whereas voice quality and verbal emotional speech belong more to the affective prosody (Bostanov and Kotchoubey, 2004; Wambacq and Jerger, 2004). It has been shown previously that 7-month-old infants can compare the emotional prosody of facial expressions and spoken words, as reflected by changes in ERP responses (Grossmann et al., 2006). In the present study, we assessed whether newborns respond to linguistic and more specifically, acoustical aspects of speech, although the presence of affective prosody may have influenced the results as well. We recorded ERFs in sleeping Finnish newborns to continuous speech (Speech condition) and singing (Singing condition) conditions in order to find an answer to this question. Both types of stimuli were developed such that they were spoken and sung in a positive and joyful mood. It may be possible that the singing condition conveyed an even happier mood than the speaking condition. It was hypothesized that, since singing contains more acoustical and affective prosodic cues than does continuous speech and since infants behaviorally respond stronger to singing than to speech (Nakata and Trehub, 2004), the response to singing would be larger than to continuous speech. Furthermore, we questioned whether the responses that were possibly obtained were indeed responses to the prosody and not a general response to stimulation itself. We tested whether reductions of prosody in continuous speech affected the magnitude of the ERF response. To this end, we introduced one condition in which the variations in the fundamental frequency were

flattened (F0 condition) and another in which the speech was low-pass filtered so that the formants were not audible (noF condition). The second manipulation was not significantly changing the affective mood of the spoken or sung signals, while the semantic content was completely diminished.

It is common to test during sleep in newborn studies, since the infants sleep most of the day. In this study, the ERFs were recorded during two sleep stages, the quiet and the active sleep (QS and AS, respectively), and the effects of sleep stage were compared, because previous research has revealed that responses during AS are very small (Pihko et al., 2004). Comparable to the English language, the Finnish language is a trochaic language. We, therefore, mainly placed the triggers to which the ERFs were recorded in the beginning of syllables starting with unvoiced stop consonants at the start of words, but also at double unvoiced stop consonants in the middle of words that were subsequently stressed (see also Section 2).

## 2. Materials and methods

### 2.1. Subjects

Auditory magnetic fields were recorded from 11 healthy newborns (6 females) 1 to 5 days after birth (mean age 2 days). The neonates were born at gestational ages of between 38 and 42 weeks. Their mean Apgar score, which is a general measure to quickly assess the health of a newborn (maximum possible score = 10), was 8.5 at 1 min. The Apgar score of one newborn was 6 at 1 min, but increased to 8 at 10 minutes. The birth weight of the neonates was between 2710 and 4100 g. The neonates were recruited from the maternity ward of the Helsinki University Central Hospital. One of the parents of each newborn signed a written informed consent prior to the study. The Ethical Committee of the Hospital District of Helsinki and Uusimaa approved the study plan.

### 2.2. Data acquisition

Magnetoencephalography (MEG) was recorded using a helmet-shaped Vectorview magnetometer (Elekta NeuroMag Oy, Helsinki) of the BioMag Laboratory, Helsinki University Central Hospital, with 306 channels. The channels are located at 102 positions uniformly over the adult head with two orthogonal planar gradiometers and one magnetometer at each location. The measuring helmet was in its supine position and the newborn was lying on his/her right side with the right hemisphere over the 'occipital' part of the helmet, enabling the recording of responses of the right hemisphere of the neonates. The right hemisphere was examined because previous studies, both with MEG and fMRI, have shown larger activation of the right compared to the left hemisphere in the processing of prosody and melody information (Patterson et al., 2002; Tervaniemi and Hugdahl, 2003).

The head position inside the MEG helmet was verified at the beginning of the recording and again later, in case the recording had to be temporarily stopped. A new recording was always started when the newborn moved significantly, as indicated by the behavioral scoring of the experimenter seated inside the recording room with a direct eye-contact to the infant. Signals from four indicator coils were used that were attached to the neonate's head at locations known in an anatomical coordinate system defined by the nasion and the preauricular points (see also [Lauronen et al., 2006](#)).

EEG and EOG recordings were simultaneously made on the scalp with silver–silver chloride disk electrodes (due to technical problems, the EEG could not be used to calculate ERPs). EEG was recorded at F4, Cz, and P4 locations. EOG was recorded from the upper left and lower right corners of the eyes. The reference electrode was placed on the right mastoid and the ground on the forehead.

Both MEG and EEG were recorded with a band-pass filter of 0.03–200 Hz and sampled continuously at 600 Hz.

### 2.3. Sleep stage determination

Our goal was to determine the sleep stages in the neonates according to the classes awake, active sleep (AS) and quiet sleep (QS) typically used in neonates. This was achieved as follows.

During the recording, one researcher sat next to the newborn to determine the sleep stages on the behavioral level. If the neonate showed closed eyes, an irregular respiration pattern, saccadic eye movements, and occasional facial twitches, AS was scored. When the eyes were closed and the respiration pattern was regular, QS was scored.

Offline, the final sleep stage was determined using the MEG, EEG, EOG, and behavioral measures of the newborns. The EEG and MEG showed low-voltage high-frequency activity during AS, whereas they showed either high-voltage low-frequency activity or trace alternants (high- and low-voltage slow waves alternating) during QS. Responses were averaged according to those sleep stages.

### 2.4. Stimuli

A Finnish Christmas carol served as stimulus material in this study. A digital recording of a female native Finnish-speaker, professional singer, was performed with a high-quality condenser microphone (Brüel & Kjær 4133) in an anechoic chamber. The carol was presented to the newborns in four different conditions, namely either being sung (Singing) by the speaker, presented as normal continuous speech (Speech), presented with the variations in the fundamental frequency being excluded (F0), or presented low-pass filtered so that the formants were not audible (noF).

The F0 stimulus was extracted from the Speech stimulus by determining the variations in the fundamental frequency of the speech by using the Praat program ([Boersma, 2001](#)),

which determines the so-called Pitch-object for the periodical parts of the speech signal. The object is generated by a precise autocorrelation method. Thereafter, the fundamental frequency was altered in Praat by using the pitch synchronous overlap-and-add (PSOLA) technique to flatten the fundamental to the median value of the Speech stimulus. This synthesis does not alter the unvoiced segments of the Speech signal. This resulted in a clearly understandable version of the speech, but with an unnaturally flat tone of voice.

The noF stimulus was acquired by simply low-pass filtering the Speech stimulus. The cut-off frequency of the filter was set at 460 Hz. This frequency was determined by iterating frequencies around the lowest first formant frequency (in /i/ in Finnish) and the highest fundamental frequency of this particular speaker. After the iteration, 460 Hz was found to be successful to remove the formant information almost totally without dropping the intensity of the stimulus too much during the highest F0 syllables. This resulted in a “hummed-like” version of the speech with no possibility to understand the semantics but still preserving the prosodic variations in F0, intensity and talking speed. See [Table 1](#) for details of the four stimulus sequences.

The triggering of the epochs for averaging brain responses to continuous speech is a challenging task. [Teder et al. \(1993\)](#) placed the triggers in their pioneering work at the beginning of words after silence starting with unvoiced stop consonants. Furthermore, [Mattys and Jusczyk \(2001\)](#) found that infants have considerably more difficulty segmenting vowel-initial than consonant-initial words from fluent speech. Therefore, in this study, we placed triggers in the beginning of syllables starting with unvoiced stop consonants in the beginning of words and at double unvoiced stop consonants in the middle of words, which are associated with a clear pause in the speech signal in the Finnish language. In addition, we placed triggers in the beginning of words and after silences. The triggers were placed by hand and cross-checked by a second researcher. Timing accuracy was  $\pm 8$  ms for starting vowels and  $\pm 4$  ms for syllables in the middle of words. See [Table 2](#) for details.

The four conditions were presented to the neonates in two sets of stimuli because of practical reasons. The newborns were presented with the singing and normal speech being alternated several times (Set 1) and they heard alternations of the F0 and noF conditions (Set 2). The two sets

Table 1  
Characteristics of the stimuli in the four conditions

	Singing	Speech	F0	noF
F0 (Hz)	176–486, mean 274, median 267	75–595, mean 199, median 186	184 (flat)	75–595, mean 199, median 186
Frequency band (Hz)	50–11,990	50–11,990	50–11,990	
Duration of stream (in s)	51	34	34	34

Table 2  
Number of triggers in the different conditions

Number of triggers in the whole stimulus after	Speech/F0/ noF	Singing
word-starting/k/	2	2
word-starting/p/	10	11
word-starting/t/	3	3
other word-starting	9	12
double-/k/	3	3
double-/p/	0	0
double-/t/	8	7
mid-word/k/(preceded by a silence)	8	2
mid-word/p/(preceded by a silence)	2	2
mid-word/t/(preceded by a silence)	11	8
Total	56	49

were introduced in counterbalanced order between the neonates. Each set was repeated eight times before the next set was started. If possible, depending on how long the neonate continued sleeping, the sets were presented again until enough data were recorded in at least the QS.

The stimuli were presented to the left ear through a plastic tube with a soft earpiece. Each set of stimuli had an intensity of 60–75 dB.

### 2.5. Procedure

First, the baby was placed in a crib and EEG and EOG electrodes were attached to the skin. Subsequently, a cloth cap was placed over the head on which four marker coils were attached to determine the position of the head with respect to the magnetometer. The position of the coils in the head coordinate system was determined with an Isotrak 3D digitizer.

Subsequently, the newborn was placed on a bed next to the MEG helmet in a magnetically shielded room (ETS; Lindgren Euroshield Oy, Euro, Finland). If necessary, the baby was fed with breast milk before the recording started. One of the researchers was inside the room with the baby to code the behavior (eye movements and breathing pattern) of the neonate. EEG, EOG, and behavioral coding were later used for off-line sleep stage analysis.

The recording was started when the newborn was sleeping and the total measurement (preparation and recording) never lasted longer than 2 h. In half the newborns, Set 1 was presented first, whereas Set 2 was first introduced in the other neonates. If necessary, the sets were repeated when both sets had been presented once.

### 2.6. Analysis

Averages for each of the four conditions were made off-line, according to the sleep stages. They were baseline corrected by taking the last 100 ms before trigger onset as baseline value. The first 800 ms after trigger onset was averaged and considered the response to a certain stimulus. Trials during which the neonate moved were rejected.

Furthermore, trials in which the sleep stage could not be determined or during which an intermediate sleep stage was present were excluded. Next, the Signal Space Separation method (Taulu et al., 2004) was performed in several newborns to remove artefacts that were coming from outside the helmet, but were recorded by the MEG instrument (e.g., heart beats and line frequency noise).

Subsequently, vector sums were calculated from the gradiometer channel pairs. The vector sums were 40 Hz low-pass filtered. Next, a mean amplitude was calculated of three (for the sleep stage comparison) and two close gradiometer pairs (for speech manipulation comparison) for each infant in order to increase signal-to-noise ratio. The gradiometer pairs were chosen individually so that they were those neighboring pairs that showed the largest responses. The vector sums were used to compare the amplitudes of the different conditions and the two sleep stages.

One predominant deflection (P1m) was found in the magnetic responses of the newborns within the range of 200–500 ms after stimulus onset. For each neonate, the latency of the maximal amplitude of this deflection was determined and the mean amplitude of 60 ms around this maximum was used for statistical analysis.

Since the data were normally distributed, the parametric analysis of variance (ANOVA) was used to statistically test the results. Only responses to which at least 125 acceptable trials were recorded within one newborn were included in the statistical analysis. In all 11 neonates, enough trials were recorded to both the Singing and Speech conditions during the QS. A Univariate ANOVA with condition (Singing vs. Speech) as within-subject factor was performed on these data. In seven out of eleven newborns (3 males and 4 females), enough trials to both Singing and Speech were recorded in both sleep stages. An ANOVA was performed on these data with condition (Singing vs. Speech) and Sleep stage (QS vs. AS) as within-subject factors. Finally, in five newborns (4 males and 1 female), a sufficient number of trials were recorded to the Speech, F0, and noF conditions during the QS. These data were statistically tested by an ANOVA with condition (speech, F0, and noF) as within-subject factor. The level of significance was set at 0.05 throughout.

## 3. Results

### 3.1. Speech vs. Singing

Fig. 1 shows a group of gradiometer channels for one newborn recorded over the right hemisphere of the newborn. Presented are the responses during the QS to the Speech and Singing conditions. The P1m response can be detected with a latency of 340 ms after stimulus onset. Fig. 2 shows a representative gradiometer channel for all newborns to the Speech and Singing conditions. The P1m over all newborns had a mean latency of 305 ms. All newborns revealed the P1m in the QS, whereas this deflection was less prominent in the AS. See Table 3 for the mean

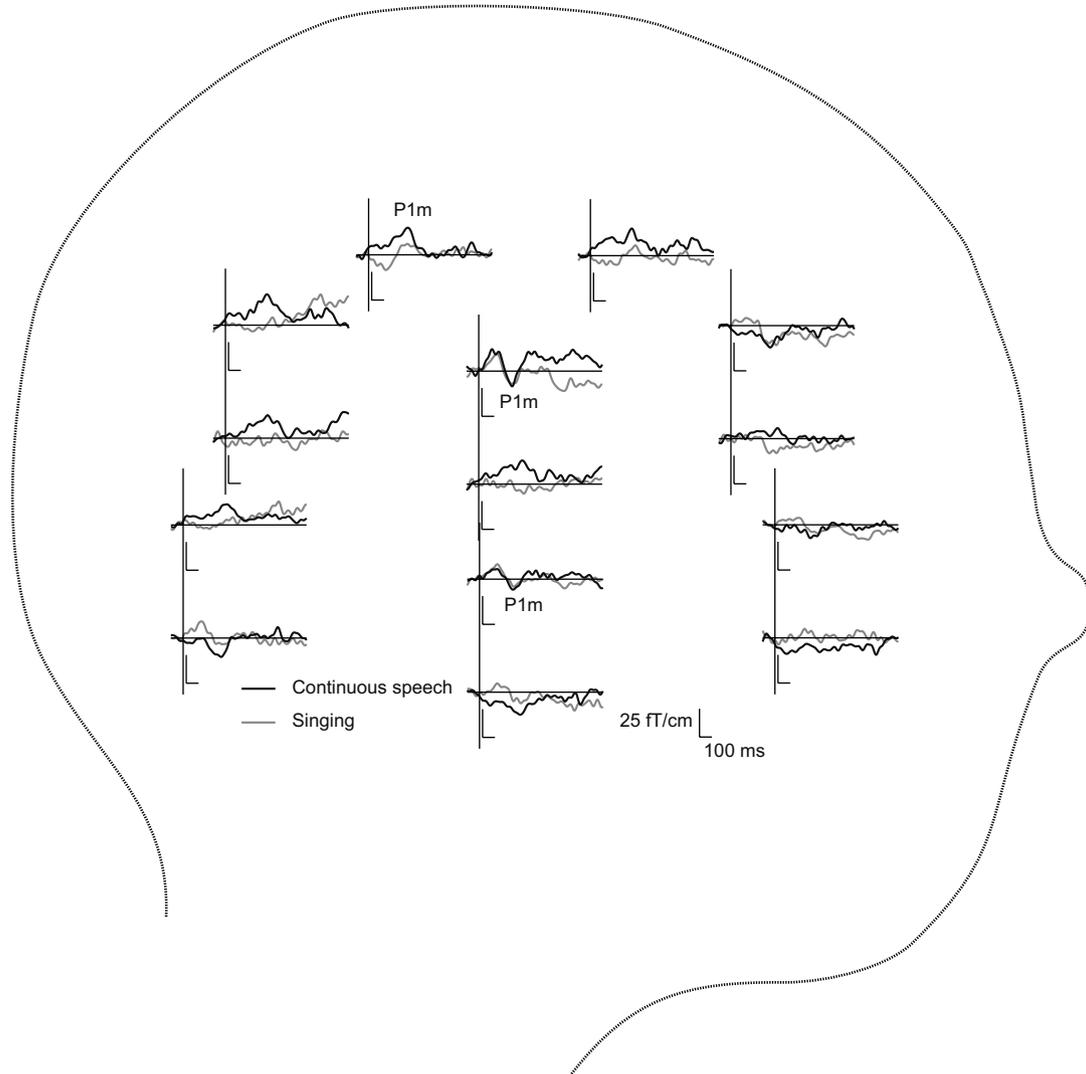


Fig. 1. A subset of gradiometer channels of one representative newborn during QS from the ‘occipital’ part of the helmet, on which the right hemisphere of this neonate was lying. The outline of the head shows the anterior–posterior array. Evoked magnetic responses to the Speech and Singing conditions. One deflection can be determined at a latency of 305 ms after word onset, pointed out as P1m.

amplitudes and standard errors of all newborns in the different conditions.

Statistical comparison of the condition (Speech vs. Singing) in 11 newborns during the QS revealed no significant differences ( $F_{1,10} = 0.44$ ;  $P > 0.52$ ; repeated measures ANOVA; sphericity assumed), as can be seen in Fig. 2. The ANOVA on condition  $\times$  sleep stage (QS vs. AS), performed on the responses of seven neonates, revealed a significant effect of Sleep stage ( $F_{1,6} = 7.82$ ;  $P = 0.03$ ), but again no effect of condition ( $F_{1,6} = 0.72$ ;  $P > 0.43$ ) and neither an interaction ( $F_{1,6} = 0.11$ ;  $P > 0.74$ ). The responses were larger during the QS than during the AS, as can also be seen in Fig. 2.

### 3.2. Natural vs. degraded speech

Fig. 3 shows a group of gradiometer channels for one newborn to the Speech, F0, and noF conditions during

the QS. Fig. 4 illustrates one representative channel for each of the five neonates in which a sufficient number of acceptable trials were recorded during Speech, F0, and noF in the QS. The mean amplitudes for each of the stimulus types are presented in Table 3. It can be seen that the P1m is most prominent in the Speech condition.

An ANOVA on condition (Speech vs. F0 vs. noF) revealed a significant effect ( $F_{2,8} = 6.40$ ;  $P = 0.02$ ). A test of within-subject contrasts revealed a linear decrease in amplitude from Speech to F0 to noF ( $F_{1,4} = 10.18$ ;  $P = 0.03$ ), indicating that the P1m was larger to Speech than to the F0 manipulation, which was again larger than the noF manipulation.

## 4. Discussion

In our study, we assessed questions related to perceiving natural, continuous speech and singing by the neonatal

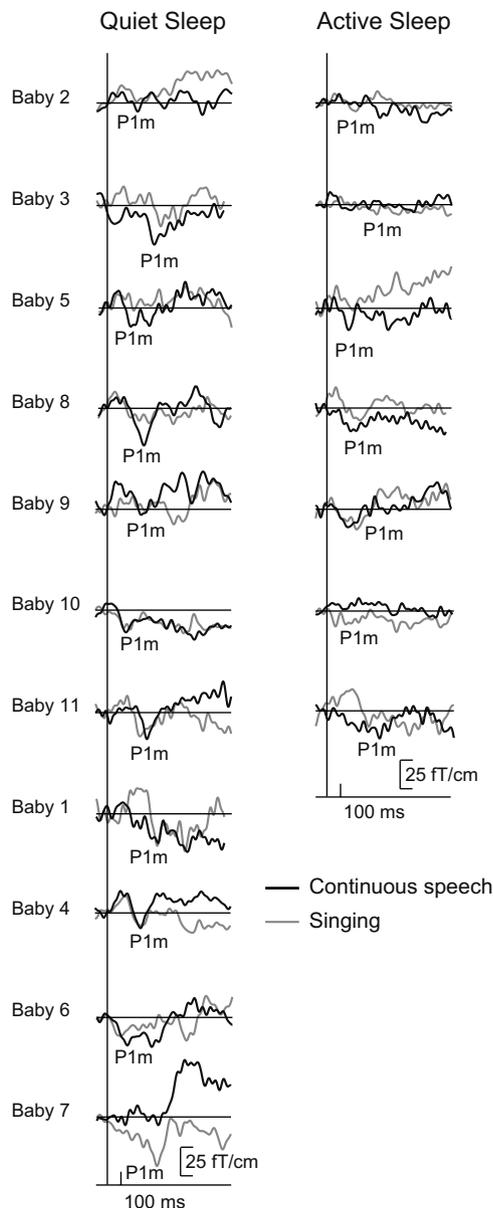


Fig. 2. Evoked magnetic fields of 11 newborns to the Speech and Singing conditions. For each infant, one representative gradiometer channel is shown (i.e., the channel with the largest general brain response). In seven out of eleven neonates, a sufficient number of epochs were recorded in both sleep stages. In the other four, only enough trials were obtained during the quiet sleep.

brain during natural sleep. Our first research question was whether valuable brain responses could be obtained to continuous speech and singing and furthermore, whether singing elicited larger responses than did speech. We found one deflection, P1m, in all newborns, which was equally large to both speech and singing. The P1m was, furthermore, larger during quiet (QS) than during active sleep (AS). Our second aim was to manipulate the speech in order to examine whether the responses elicited in the newborns were due to the prosody and not to stimulation per se. The amplitude of the P1m decreased when the natural variations of the fundamental frequency were excluded and further

Table 3

Mean amplitudes (fT/cm) with standard error (in parentheses) of the vector sums of deflection P1m in response to auditory stimulation for the Singing, Speech, F0, and noF conditions

	n	Condition		
		Singing vs. Speech	Singing vs. Speech × QS vs. AS	Speech vs. F0 vs. noF
Speech QS	11	19.72 (1.87)	x	x
Singing QS	11	23.57 (4.39)	x	x
Speech QS	7	x	18.63 (1.80)	x
Singing QS	7	x	21.60 (6.20)	x
Speech AS	7	x	12.89 (6.04)	x
Singing AS	7	x	13.66 (9.54)	x
Speech QS	5	x	x	21.30 (2.35)
F0 QS	5	x	x	16.55 (2.73)
noF QS	5	x	x	12.30 (1.36)

Note that the responses are all positive, due to the calculation method used for computing vector sums. Thus, these amplitudes do not provide information about the polarity of the signal.

decreased when low-pass filtering the signal and, thereby, making the formants inaudible, even though several prosodic features still remained.

One important point to note is the effect of the sleep stage on the brain responses. Effects of sleep stage have not been found in all studies. For instance, [Martynova et al. \(2003\)](#) studied whether the mismatch response to vowel changes is larger to QS than to AS in newborns. They, however, failed to find a significant effect. [Pihko et al. \(2004\)](#), on the other hand, showed that the standard response to a syllable consisting of a consonant and a vowel was larger during QS than during AS. The stimuli used in their study were more complex than the vowels used by [Martynova et al. \(2003\)](#). The stimuli used in the current study were even more complex and showed a clear sleep stage effect. It is, therefore, important to separate the responses to the different sleep stages when examining responses to complex speech in sleeping infants, at least in newborns, and to focus on the QS, where clear responses can be obtained.

Recently, [Kooiman et al. \(2005\)](#) recorded ERPs to continuous speech in awake 10-month-old infants. In each block of stimuli, the infants were first familiarized with a two-syllable word, which was followed by the presentation of several sentences that did or did not include that word. The infants revealed a deflection around 350–500 ms to word onsets in the continuous speech, which significantly differed between the familiar and unfamiliar words. We also found one clear deflection with a similar latency, only in sleeping newborns. As already mentioned, the infant ERP seems to be relatively stable throughout the first year ([Kushnerenko et al., 2002b](#)). The present results and those of Kooiman and her colleagues confirm this notion.

The main aim of our study was to examine whether the possibility to respond to linguistic/acoustical prosody is already present at birth, or not. [Thiessen and Saffran](#)

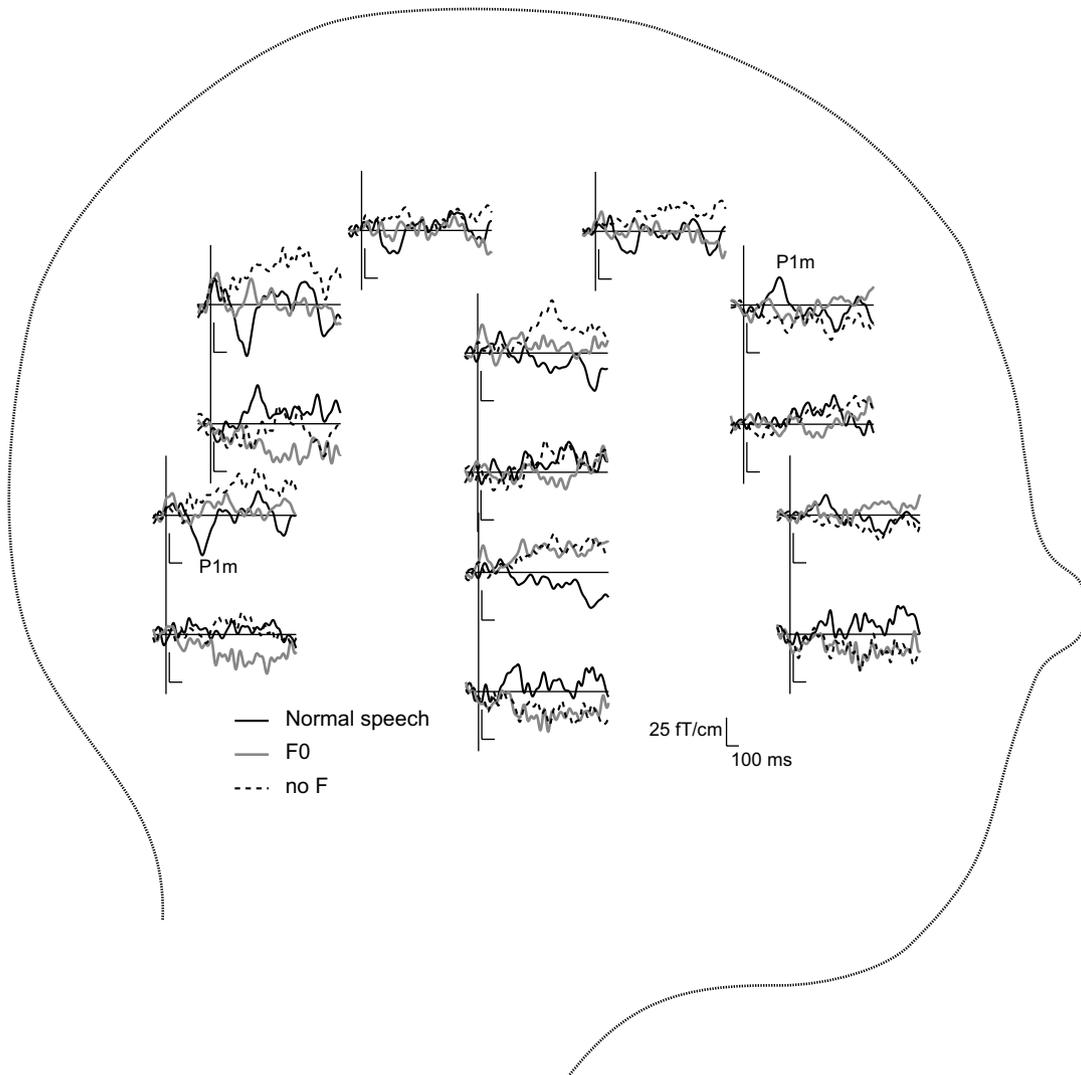


Fig. 3. A subset of gradiometer channels of one representative newborn to the Speech, F0, and noF conditions. Note that a clear P1m can be observed to the Speech, whereas it is dramatically decreased to the other two conditions.

(2003) concluded that infants at different ages may use different segmentation strategies, since statistical cues were more important in language acquisition at the age of 7 months, whereas prosody only becomes important later, at the age of 9 months. We have shown that newborns do respond to trochaic items already at birth. After all, a clear P1m response could be obtained in all newborns during continuous speech and singing. Furthermore, when decreasing the amount of prosody in the continuous speech, the amplitude of P1m dramatically decreased as well. These results suggest that newborns indeed have the opportunity to use prosody, which is in accordance with the findings of [Pena et al. \(2003\)](#), who found brain activation to speech stimuli in healthy newborns. However, the current results do not necessarily imply that newborns have immediately started acquiring language by using prosody. The results show that the auditory cortex at the newborn age is capable of extracting information from the prosodic features and automatically responds to certain prosodic cues in the auditory domain.

One remark must be made. We hypothesized that singing would elicit larger responses than continuous speech, because singing may include even stronger cues than speech, not only acoustically, but also affectively. Also, in singing, the constancy of the rhythmic patterns might help the auditory cortex to extract the beginnings of phrases. However, the amplitudes of P1m did not significantly differ between the two stimulation types. [Thiessen and Saffran \(2004\)](#) studied in 9-month-old infants whether one single stress cue, spectral tilt, is treated as a strong indicator for word boundaries in the same way as multiple stress cues are strong indicators, shown in one of their previous studies ([Thiessen and Saffran, 2003](#)). The results showed that this was indeed the case. [Thiessen and Saffran \(2004\)](#), therefore, concluded that a single cue to stress has the same effect as multiple cues to stress, at least in their paradigm. Our results showing similar responses to continuous speech and singing might also be explained by this statement. As long as stress cues are available in a syllable, newborns show a brain response to them. Thus, even though the

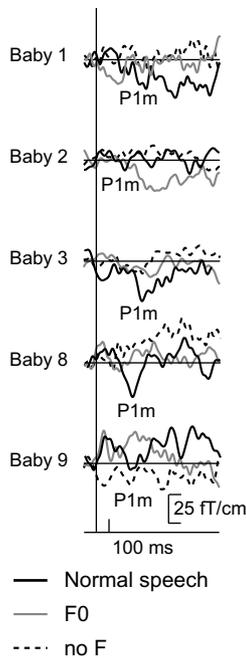


Fig. 4. Evoked magnetic fields of five neonates to the Speech, F0, and noF conditions. For each infant, one representative gradiometer channel is shown. Note again that the P1m can be clearly seen to the speech, but less to the F0 and noF conditions.

singing condition may have contained more affective prosody as compared to the continuous speech condition, it did not result in larger responses to singing due to some kind of a ceiling effect.

In adults, differential responses to affective or semantic stimuli have been found (Wambacq and Jerger, 2004), as well as differences in responses to various emotional features (Bostanov and Kotchoubey, 2004). Affective prosody is, thus, a very important aspect of language processing. As noted above, we did not find any differences between the singing and speech, which could have been related to the fact that a ceiling effect was reached or to a small sample size. More research is needed to find out whether newborn infants possess the ability to distinguish between stimuli that contain variable amounts of emotional prosody.

The F0 condition also contained some prosody, mainly linguistic. The pitch differences were excluded from the speech, but the first syllable of each word was still stressed, as indicated by the intensity and slightly also by the duration of the syllable. This differential effect of Thiessen and Saffran (2004) and the current study may be due to the type of response that is measured. By using ERPs or ERFs, a more precise measure is taken than by using the head turning procedure. It is, therefore, possible that the amount of stress cues affects the responses of newborns. There might, however, be a ceiling effect: with a certain amount of prosody, no changes will be found anymore.

The question that is still open is which aspect of prosody is most important in processing language. According to the current results, it seems that ‘syllable stress’ is less essential than the F0 variations, because the response was decreased

when only syllable stress was the major prosodic cue in the speech. We did not find any differences between the singing and normal speech conditions, thus the fact that singing also contained affective prosody to some extent did not seem to be crucial, at least not with the current setup. Possibly one prosodic feature that was shared in the two conditions is crucial, such as pitch of the stimulus. More research will be necessary, however, to determine the importance of each of the prosodic cues.

It is somewhat difficult to interpret why the responses were even more diminished when the formant frequencies were filtered out. One simple explanation is related to the general intensity level of the stimulus. We measured the intensities of the stimuli by using the A-dB-filter, which mimics adult hearing. This means that for an adult auditory system, the acoustic energy was the same in the different stimulus types. In neonates we cannot know exactly, however, how large the effect is of filtering the higher part of the acoustic spectrum out, since the auditory system is not yet fully developed. The energy in the remaining lower part might not be sufficient in neonates to elicit the P1m response as strongly as in the other stimuli in which a wide range of frequencies were present. For this reason, it is very difficult to draw strong conclusions on the result that the P1m was very small for the speech signal without formants.

In conclusion, the present study shows that the use of natural, continuous speech ERF recordings of neonates is feasible, and that newborns are sensitive to certain prosodic aspects of auditory input.

## Acknowledgements

This work, as part of the European Science Foundation EUROCORES Programme OMLL, was supported by funds from the Academy of Finland and the EC Sixth Framework Programme under Contract No. ERAS-CT-2003-980409.

## References

- Boersma P. Praat, a system for doing phonetics by computer. *Glott Int* 2001;5:341–5.
- Bostanov V, Kotchoubey B. Recognition of affective prosody: continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology* 2004;41:259–68.
- Ceponiene R, Kushnerenko E, Fellman V, Renlund M, Suominen K, Näätänen R. Event-related potential features indexing central auditory discrimination by newborns. *Cogn Brain Res* 2002;13:101–13.
- Cheour M, Alho K, Čeponiene R, Reinikainen K, Sainio K, Phjavuori M, et al. Maturation of mismatch negativity in infants. *Int J Psychophysiol* 1998;29:217–26.
- Cheour M, Kushnerenko E, Čeponiene R, Fellman V, Näätänen R. Electric brain responses obtained from newborn infants to changes in duration in complex harmonic tones. *Dev Neuropsychol* 2002a;22:471–9.
- Cheour M, Martynova O, Näätänen R, Erkola R, Sillanpää M, Kero P, et al. Speech sounds learned by sleeping newborns. *Nature* 2002b;45:599–600.
- Cheour M, Imada T, Taulu S, Ahonen A, Salonen J, Kuhl P. Magnetoencephalography is feasible for infant assessment of auditory discrimination. *Exp Neurol* 2004;190:S44–51.

- Dehaene-Lambertz G, Pena M. Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport* 1998;12:3155–8.
- Ellingson RJ, Danahy T, Nelson B, Lathrop GH. Variability of auditory evoked potentials in human newborns. *Electroenceph Clin Neurophysiol* 1974;36:155–62.
- Grossmann T, Striano T, Friederici AD. Crossmodal integration of emotional information from face and voice in the infant brain. *Dev Sci* 2006;3:309–15.
- Huotilainen M, Kujala A, Hotakainen M, Shestakova A, Kushnerenko E, Parkkonen L, et al. Auditory magnetic responses of healthy newborns. *Neuroreport* 2003;14:1871–3.
- Johnson EK, Jusczyk PW. Word segmentation by 8-month-olds: when speech cues count more than statistics. *J Mem Lang* 2001;44:548–67.
- Jusczyk PW, Houston DM, Newsome M. The beginnings of word segmentation in english-learning infants. *Cogn Psychol* 1999;39:159–207.
- Kooiman V, Hagoort P, Cutler A. Electrophysiological evidence for prelinguistic infant's word recognition in continuous speech. *Cogn Brain Res* 2005;24:109–16.
- Kuhl PK. Early language acquisition: cracking the speech code. *Nat Rev Neurosci* 2004;5:831–43.
- Kujala A, Huotilainen M, Hotakainen M, Lennes M, Parkkonen L, Fellman V, et al. Speech-sound discrimination in neonates as measured with MEG. *Neuroreport* 2004;15:2089–92.
- Kurtzberg D, Hilpert PL, Kreuzer JA, Vaughan Jr HG. Differential maturation of cortical auditory evoked potentials to speech sounds in normal fullterm and very low-birth weight infants. *Dev Med Child Neurol* 1984;26:466–75.
- Kushnerenko E, Čeponiene R, Balan P, Fellman V, Näätänen R. Maturation of the auditory change detection response in infants: a longitudinal ERP study. *Neuroreport* 2002a;13:1843–8.
- Kushnerenko E, Čeponiene R, Balan P, Fellman V, Huotilainen M, Näätänen R. Maturation of the auditory event-related potentials during the first year of life. *Neuroreport* 2002b;13:47–51.
- Lauronen L, Nevalainen P, Wikström H, Parkkonen L, Okada Y, Pihko E. Immaturity of somatosensory cortical processing in human newborns. *Neuroimage* 2006;33:195–203.
- Lengle JM, Chen M, Wakai RT. Improved neuromagnetic detection of fetal and neonatal auditory evoked responses. *Clin Neurophysiol* 2001;112:785–92.
- Martynova O, Kirjavainen J, Cheour M. Mismatch negativity and late discriminative negativity in sleeping human newborns. *Neurosci Lett* 2003;340:75–8.
- Mattys SL, Jusczyk PW. Do infants segment words or recurring contiguous patterns? *J Exp Psychol* 2001;27:644–55.
- Näätänen R, Winkler I. The concept of auditory stimulus representation in cognitive neuroscience. *Psychol Bull* 1999;125:826–59.
- Nakata T, Trehub SE. Infant's responsiveness to maternal speech and singing. *Infant Behav Dev* 2004;27:455–64.
- Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD. The processing of temporal pitch and melody information in auditory cortex. *Neuron* 2002;36:767–76.
- Pena M, Maki A, Kovacic D, Gehaene-Lambertz G, Koizumi H, Bouquet F, et al. Sounds and silence: an optical topography study of language recognition at birth. *PNAS* 2003;100:11702–5.
- Pihan H. Affective and linguistic processing of speech prosody: DV potential studies. *Prog Brain Res* 2006;156:269–84.
- Pihko E, Leppänen PHT, Eklund KM, Cheour M, Guttorm TK, Lyytinen H. Cortical responses of infants with and without a genetic risk for dyslexia: I. Age effects. *Neuroreport* 1999;10:901–5.
- Pihko E, Sambeth A, Leppänen P, Okada Y, Lauronen L. Auditory evoked magnetic fields to speech stimulus in newborns – effect of sleep stages. *Neurol Clin Neurophysiol* 2004;6:1–5.
- Saffran JR. Words in a sea of sounds: the output of infant statistical learning. *Cognition* 2001;81:149–69.
- Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science* 1996;274:1926–8.
- Sambeth A, Huotilainen M, Kushnerenko E, Fellman V, Pihko E. Newborns discriminate novel from harmonic sounds: a study using magnetoencephalography. *Clin Neurophysiol* 2006;117:496–503.
- Taulu S, Simola J, Kajola M. MEG recordings of DC fields using the signal space separation method (SSS). *Neurol Clin Neurophysiol* 2004;35:1–4.
- Teder W, Kujala T, Näätänen R. Selection of speech messages in free-field listening. *Neuroreport* 1993;5:307–9.
- Tervaniemi M, Hugdahl K. Lateralization of auditory-cortex functions. *Brain Res Rev* 2003;43:231–46.
- Thiessen ED, Saffran JR. When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Dev Psychol* 2003;39:706–16.
- Thiessen ED, Saffran JR. Spectral tilt as a cue to word segmentation in infancy and adulthood. *Percept Psychophys* 2004;66:779–91.
- Wambacq IJA, Jerger JF. Processing of affective prosody and lexical-semantics in spoken utterances as differentiated by event-related potentials. *Cogn Brain Res* 2004;20:427–37.
- Weber C, Hahne A, Friedrich M, Friederici AD. Discrimination of word stress in early infant perception: electrophysiological evidence. *Cogn Brain Res* 2004;18:149–61.